

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Lobel, Peter  
Sleat, David E.

(ii) TITLE OF INVENTION: NOVEL HUMAN LYSOSOMAL PROTEIN AND  
METHODS OF ITS USE

(iii) NUMBER OF SEQUENCES: 12

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: David A. Jackson, Esq.  
(B) STREET: 411 Hackensack Ave, Continental Plaza, 4th  
Floor  
(C) CITY: Hackensack  
(D) STATE: New Jersey  
(E) COUNTRY: USA  
(F) ZIP: 07601

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US  
(B) FILING DATE:  
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Jackson Esq., David A.  
(B) REGISTRATION NUMBER: 26,742  
(C) REFERENCE/DOCKET NUMBER: 601-1-077

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 201-487-5800  
(B) TELEFAX: 201-343-1684

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3487 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGCGGAAGGG CAGAATGGGA CTCCAAGCCT GCCTCCTAGG GCTCTTGCC CTCATCCTCT

60

CTGGCAAATG CAGTTACAGC CCGGAGCCCG ACCAGCGGAG GACGCTGCC CCAGGGCTGGG

120

TGTCCCTGGG	CCGTGCGGAC	CCTGAGGAAG	AGCTGAGTCT	CACCTTGCC	CTGAGACAGC	180
AGAATGTGGA	AAGACTCTCG	GAGCTGGTGC	AGGCTGTGTC	GGATCCCAGC	TCTCCTCAAT	240
ACGGAAAATA	CCTGACCCCTA	GAGAATGTGG	CTGATCTGGT	GAGGCCATCC	CCACTGACCC	300
TCCACACGGT	GCAAAAATGG	CTCTTGGCAG	CCGGAGCCCA	GAAGTGCAT	TCTGTGATCA	360
CACAGGACTT	TCTGACTTGC	TGGCTGAGCA	TCCGACAAGC	AGAGCTGCTG	CTCCCTGGGG	420
CTGAGTTCA	TCACTATGTG	GGAGGACCTA	CGGAAACCCA	TGTTGTAAGG	TCCCCACATC	480
CCTACCAGCT	TCCACAGGCC	TTGGCCCCCC	ATGTGGACTT	TGTGGGGGA	CTGCACCATT	540
TTCCCCAAC	ATCATCCCTG	AGGCAACGTC	CTGAGCCGA	GGTGACAGGG	ACTGTAGGCC	600
TGCATCTGGG	GGTAACCCCC	TCTGTGATCC	GTAAGCGATA	CAACTTGACC	TCACAAGACG	660
TGGGCTCTGG	CACCAGCAAT	AACAGCCAAG	CCTGTGCCA	GTTCTGGAG	CAGTATTCC	720
ATGACTCAGA	CCTGGCTCAG	TTCATGCGCC	TCTTCGGTGG	CAACTTTGCA	CATCAGGCAT	780
CAGTAGCCCG	TGTGGTTGGA	CAACAGGGCC	GGGGCCGGGC	CGGGATTGAG	GCCAGTCTAG	840
ATGTGCAGTA	CCTGATGAGT	GCTGGTGCCA	ACATCTCCAC	CTGGGTCTAC	AGTAGCCCTG	900
GCCGGCATGA	GGGACAGGAG	CCCTTCCTGC	AGTGGCTCAT	GCTGCTCAGT	AATGAGTCAG	960
CCCTGCCACA	TGTGCATACT	GTGAGCTATG	GAGATGATGA	GGACTCCCTC	AGCAGCGCCT	1020
ACATCCAGCG	GGTCAACACT	GAGCTCATGA	AGGCTGCTGC	TCGGGGTCTC	ACCCTGCTCT	1080
TCGCCTCAGG	TGACAGTGGG	GCCGGGTGTT	GGTCTGTCTC	TGGAAGACAC	CAGTTCCGCC	1140
CTACCTTCCC	TGCCTCCAGC	CCCTATQTCA	CCACAGTGGG	AGGCACATCC	TTCCAGGAAC	1200
CTTTCTCAT	CACAAATGAA	ATTGTTGACT	ATATCAGTGG	TGGTGGCTTC	AGCAATGTGT	1260
TCCCACGGCC	TTCATACCAAG	GAGGAAGCTG	TAACGAAGTT	CCTGAGCTCT	AGCCCCCACC	1320
TGCCACCATC	CAGTTACTTC	AATGCCAGTG	GCCGTGCCTA	CCCAGATGTG	GCTGCACTTT	1380
CTGATGGCTA	CTGGGTGGTC	AGCAACAGAG	TGCCCATTCC	ATGGGTGTCC	GGAACCTCGG	1440
CCTCTACTCC	AGTGTGTTGGG	GGGATCCTAT	CCTTGATCAA	TGAGCACAGG	ATCCTTAGTG	1500
GCCGCCCCCC	TCTTGGCTTT	CTCAACCCAA	GGCTCTACCA	GCAGCATGGG	GCAGGACTCT	1560
TTGATGTAAC	CCGTGGCTGC	CATGAGTCCT	GTCTGGATGA	AGAGGTAGAG	GGCCAGGGTT	1620
TCTGCTCTGG	TCCTGGCTGG	GATCCTGTAA	CAGGCTGGGG	AACACCCAAAC	TTCCCAGCTT	1680
TGCTGAAGAC	TCTACTCAAC	CCCTGACCCCT	TTCCTATCAG	GAGAGATGGC	TTGTCCCCCTG	1740
CCCTGAAGCT	GGCAGTTCAAG	TCCCTTATTC	TGCCCTGTTG	GAAGCCCTGC	TGAACCCCTCA	1800
ACTATTGACT	GCTGCAGACA	GCTTATCTCC	CTAACCCCTGA	AATGCTGTGA	GCTTGACTTG	1860
ACTCCCAACC	CTACCATGCT	CCATCATACT	CAGGTCTCCC	TACTCCTGCC	TTAGATTCCCT	1920
CAATAAGATG	CTGTAACTAG	CATTTTTTGA	ATGCCCTCTCC	CTCCGCATCT	CATCTTTCTC	1980

TTTTCAATCA GGCTTTCCA AAGGGTTGTA TACAGACTCT GTGCACTATT TCACTTGATA	2040
TTCAATTCCCC AATTCACTGC AAGGAGACCT CTACTGTCAC CGTTTACTCT TTCCTACCCCT	2100
GACATCCAGA AACAAATGGCC TCCAGTGCAT ACTTCTCAAT CTTTGCTTTA TGGCCTTCC	2160
ATCATAGTTG CCCACTCCCT CTCCTTACTT AGCTTCCAGG TCTTAACCTTC TCTGACTACT	2220
CTTGTCTTCC TCTCTCATCA ATTTCTGCTT CTTCATGGAA TGCTGACCTT CATTGCTCCA	2280
TTTGTAGATT TTTGCTCTTC TCAGTTACT CATTGTCCCC TGGAACAAAT CACTGACATC	2340
TACAACCATT ACCATCTCAC TAAATAAGAC TTTCTATCCA ATAATGATTG ATACCTCAAA	2400
TGTAAGATGC GTGATACTCA ACATTTCATC GTCCACCTTC CCAACCCAA ACAATTCCAT	2460
CTCGTTTCTT CTTGGTAAAT GATGCTATGC TTTTCCAAC CAAGCCAGAA ACCTGTGTCA	2520
TCTTTTCACC CCACCTTCAA TCAACAAGTC CTCAATCAAC AAGTCCTACT GACTGCACAT	2580
CTTAAATATA TCTTTATCAG TCCACAAGTC CTTCCAATTAA TATTTCCCAA GTATATCTAG	2640
AACTTATCCA CTTATATCCC CACTGCTACT ACCTTAGTTT AGGGCTATAT TCTCTTGAAA	2700
AAAAGTGTCC TTACTTCCTG CCAATCCCCA AGTCATCTTC CAGAGTAAAA TGCAAATCCC	2760
ATCAGGCCAC TTGGATGAAA ACCCTTCAAG GATTACTGGA TAGAATTCAAG GCTTTCCCT	2820
CCASCCCCCA ATCATAGCTC ACAAAACCTTC CTTGCTATTG GTTCTTAAGT AAAAAATCAT	2880
TTTCCTCCT CCCTCCCCAA ACCCCAAGGA ACTCTCACTC TTGCTCAAGC TGTTCCGTCC	2940
CCTTACCACC CCTGATACAA CTGCCAGGTT AATTCCAGA ATTCTTGCAA GACTCAGTTC	3000
AGAAGTCACC TTCTTCGTG AATGTTTGAA TTCCCTGAGG CTACTTTATT TTGGTATGGC	3060
TGAAAAATCC TAGATTTCT AAACAAAACC TGTTGAATC TTGGTTCTGA TATGGACTAG	3120
GAGAGAGACT GGGTCAAGTA AGCTTATCTC CCTGAGGCTG TTCCCTCGTC TGTTAAGTGT	3180
GAATATCAAT ACCTGCCCTT CATAATCACC AGGGAATAAA GTGGAATAAT GTTGATAACA	3240
GTGCTTGGCA CCTGGAAGTA GGTGGCAGAT GTTAACGCC TTCCCTCCTT GCACTGCGCC	3300
CCCTGTGCCT ACCTCTAGCA TTGTAACGAC CACATAGTAT TGAAATGCC AGTTTACTTG	3360
TCTGCCTTCC TTTCCAAGAC CGTTGGTGCC TAGAGGACTA GAATCGTGTCTATTTAACT	3420
TTGTGTTCCC AGGTCTAGC TCAGGAGTTG GCAAATAAGA ATTAAATGTC TGCTACACCG	3480
AAACAAA	3487

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2520 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGCGGAAGGG CAGAATGGGA CTCCAAGCCT GCCTCCTAGG GCTCTTGCC CTCATCCTCT	60
CTGGCAAATG CAGTTACAGC CCGGAGCCCG ACCAGCGGAG GACGCTGCC CCAGGCTGGG	120
TGTCCCTGGG CCGTGGGAC CCTGAGGAAG AGCTGAGTCT CACCTTGCC CTGAGACAGC	180
AGAATGTGGA AAGACTCTCG GAGCTGGTGC AGGCTGTGTC GGATCCCAGC TCTCCTCAAT	240
ACGGAAAATA CCTGACCCCTA GAGAATGTGG CTGATCTGGT GAGGCCATCC CCACTGACCC	300
TCCACACGGT GCAAAAATGG CTCTTGGCAG CCGGAGCCCA GAAGTGCCAT TCTGTGATCA	360
CACAGGACTT TCTGACTTGC TGGCTGAGCA TCCGACAAGC AGAGCTGCTG CTCCCTGGGG	420
CTGAGTTCA TCACTATGTG GGAGGACCTA CGGAAACCCA TGTTGTAAGG TCCCCACATC	480
CCTACCAGCT TCCACAGGCC TTGGCCCCC ATGTGGACTT TGTGGGGGA CTGCACCATT	540
TTCCCCAAC ATCATCCCTG AGGCAACGTC CTGAGCCGCA GGTGACAGGG ACTGTAGGCC	600
TGCATCTGGG GGTAACCCCC TCTGTGATCC GTAAGCGATA CAACTTGACC TCACAAGACG	660
TGGGCTCTGG CACCAGCAAT AACAGCCAAG CCTGTGCCA GTTCCTGGAG CAGTATTTC	720
ATGACTCAGA CCTGGCTCAG TTCATGCGCC TCTTCGGTGG CAACTTGCA CATCAGGCAT	780
CAGTAGCCCCG TGTGGTTGGA CAACAGGGCC GGGGCCGGGC CGGGATTGAG GCCAGTCTAG	840
ATGTGCAGTA CCTGATGAGT GCTGGTGCAC ACATCTCCAC CTGGGTCTAC AGTAGCCCTG	900
GCCGGCATGA GGGACAGGAG CCCTTCCTGC AGTGGCTCAT GCTGCTCACT AATGAGTCAG	960
CCCTGCCACA TGTGCATACT GTGAGCTATG GAGATGATGA GGACTCCCTC AGCAGCGCCT	1020
ACATCCAGCG GGTCAACACT GAGCTCATGA AGGCTGCTGC TCGGGTCTC ACCCTGCTCT	1080
TGCCCTCAGG TGACAGTGGG GCCGGGTGTT GGTCTGTCTC TGGAAAGACAC CAGTTCCGCC	1140
CTACCTTCCC TGCCTCCAGC CCCTATGTCA CCACAGTGGG AGGCACATCC TTCCAGGAAC	1200
CTTTCCCTCAT CACAAATGAA ATTGTTGACT ATATCAGTGG TGGTGGCTTC AGCAATGTGT	1260
TCCCACGGCC TTCATACCAAG GAGGAAGCTG TAACGAAGTT CCTGAGCTCT AGCCCCCACC	1320
TGCCACCATC CAGTTACTTC AATGCCAGTG GCCGTGCCTA CCCAGATGTG GCTGCACTTT	1380
CTGATGGCTA CTGGGTGGTC AGAACACAGAG TGCCCATTC ATGGGTGTCC GGAACCTCGG	1440
CCTCTACTCC AGTGTGTTGGG GGGATCCTAT CCTTGATCAA TGAGCACAGG ATCCTTAGTG	1500
GCCGCCCCCCC TCTTGGCTTT CTCAACCCAA GGCTCTACCA GCAGCATGGG GCAGGACTCT	1560
TTGATGTAAC CCGTGGCTGC CATGAGTCCT GTCTGGATGA AGAGGTAGAG GGCCAGGGTT	1620
TCTGCTCTGG TCCTGGCTGG GATCCTGTAA CAGGCTGGGG AACACCCAAAC TTCCCAGCTT	1680
TGCTGAAGAC TCTACTCAAC CCCTGACCCCT TTCCTATCAG GAGAGATGGC TTGTCCCCTG	1740

CCCTGAAGCT	GGCAGTTCA	TCCCTTATT	TGCCCTGTT	GAAGCCCTGC	TGAACCCTCA	1800
ACTATTGACT	GCTGCAGACA	GCTTATCTCC	CTAACCCCTGA	AATGCTGTGA	GCTTGACTTG	1860
ACTCCCAACC	CTACCATGCT	CCATCATACT	CAGGTCTCCC	TACTCCTGCC	TTAGATTCC	1920
CAATAAGATG	CTGTAACTAG	CATTTTTGA	ATGCCTCTCC	CTCCGCATCT	CATCTTCTC	1980
TTTTCAATCA	GGCTTTCCA	AAGGGTTGTA	TACAGACTCT	GTGCACTATT	TCACTTGATA	2040
TTCATTCCCC	AATTCACTGC	AAGGAGACCT	CTACTGTCAC	CGTTTACTCT	TTCCCTACCC	2100
GACATCCAGA	AACAATGGCC	TCCAGTGCAT	ACTTCTCAAT	CTTGCTTTA	TGGCCTTTCC	2160
ATCATAGTTG	CCOACTCCCT	CTCCTTACTT	AGCTTCCAGG	TCTTAACCTTC	TCTGACTACT	2220
CTTGTCTTCC	TCTCTCATCA	ATTTCTGCTT	CTTCATGGAA	TGCTGACCTT	CATTGCTCCA	2280
TTTGTAGATT	TTTGCTCTTC	TCAGTTACT	CATTGTCCCC	TGGAACAAAT	CACTGACATC	2340
TACAACCATT	ACCATCTCAC	TAAATAAGAC	TTTCTATCCA	ATAATGATTG	ATACCTAAA	2400
TGTAAGATGC	GTGATACTCA	ACATTCATC	GTCCACCTTC	CCAACCCCAA	ACAATTCCAT	2460
CTCGTTCTT	CTTGGTAAAT	GATGCTATGC	TTTTCCAAC	CAAAAAAAAAA	AAAAAAAAAA	2520

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 563 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met	Gly	Leu	Gln	Ala	Cys	Leu	Leu	Gly	Leu	Phe	Ala	Leu	Ile	Leu	Ser
1					5				10					15	
Gly	Lys	Cys	Ser	Tyr	Ser	Pro	Glu	Pro	Asp	Gln	Arg	Arg	Thr	Leu	Pro
				20				25					30		
Pro	Gly	Trp	Val	Ser	Leu	Gly	Arg	Ala	Asp	Pro	Glu	Glu	Glu	Leu	Ser
					35			40					45		
Leu	Thr	Phe	Ala	Leu	Arg	Gln	Gln	Asn	Val	Glu	Arg	Leu	Ser	Glu	Leu
					50			55			60				
Val	Gln	Ala	Val	Ser	Asp	Pro	Ser	Ser	Pro	Gln	Tyr	Gly	Lys	Tyr	Leu
					65			70			75			80	
Thr	Leu	Glu	Asn	Val	Ala	Asp	Leu	Val	Arg	Pro	Ser	Pro	Leu	Thr	Leu
					85				90					95	

His Thr Val Gln Lys Trp Leu Leu Ala Ala Gly Ala Gln Lys Cys His  
100 105 110

Ser Val Ile Thr Gln Asp Phe Leu Thr Cys Trp Leu Ser Ile Arg Gln  
115 120 125

Ala Glu Leu Leu Leu Pro Gly Ala Glu Phe His His Tyr Val Gly Gly  
130 135 140

Pro Thr Glu Thr His Val Val Arg Ser Pro His Pro Tyr Gln Leu Pro  
145 150 155 160

Gln Ala Leu Ala Pro His Val Asp Phe Val Gly Gly Leu His His Phe  
165 170 175

Pro Pro Thr Ser Ser Leu Arg Gln Arg Pro Glu Pro Gln Val Thr Gly  
180 185 190

Thr Val Gly Leu His Leu Gly Val Thr Pro Ser Val Ile Arg Lys Arg  
195 200 205

Tyr Asn Leu Thr Ser Gln Asp Val Gly Ser Gly Thr Ser Asn Asn Ser  
210 215 220

Gln Ala Cys Ala Gln Phe Leu Glu Gln Tyr Phe His Asp Ser Asp Leu  
225 230 235 240

Ala Gln Phe Met Arg Leu Phe Gly Gly Asn Phe Ala His Gln Ala Ser  
245 250 255

Val Ala Arg Val Val Gly Gln Gln Gly Arg Gly Arg Ala Gly Ile Glu  
260 265 270

Ala Ser Leu Asp Val Gln Tyr Leu Met Ser Ala Gly Ala Asn Ile Ser  
275 280 285

Thr Trp Val Tyr Ser Ser Pro Gly Arg His Glu Gly Gln Glu Pro Phe  
290 295 300

Leu Gln Trp Leu Met Leu Leu Ser Asn Glu Ser Ala Leu Pro His Val  
305 310 315 320

His Thr Val Ser Tyr Gly Asp Asp Glu Asp Ser Leu Ser Ser Ala Tyr  
325 330 335

Ile Gln Arg Val Asn Thr Glu Leu Met Lys Ala Ala Ala Arg Gly Leu  
340 345 350

Thr Leu Leu Phe Ala Ser Gly Asp Ser Gly Ala Gly Cys Trp Ser Val  
355 360 365

Ser Gly Arg His Gln Phe Arg Pro Thr Phe Pro Ala Ser Ser Pro Tyr  
370 375 380

Val Thr Thr Val Gly Gly Thr Ser Phe Gln Glu Pro Phe Leu Ile Thr  
385 390 395 400

Asn Glu Ile Val Asp Tyr Ile Ser Gly Gly Phe Ser Asn Val Phe  
405 410 415

Pro Arg Pro Ser Tyr Gln Glu Ala Val Thr Lys Phe Leu Ser Ser  
420 425 430

Ser Pro His Leu Pro Pro Ser Ser Tyr Phe Asn Ala Ser Gly Arg Ala  
 435 440 445  
 Tyr Pro Asp Val Ala Ala Leu Ser Asp Gly Tyr Trp Val Val Ser Asn  
 450 455 460  
 Arg Val Pro Ile Pro Trp Val Ser Gly Thr Ser Ala Ser Thr Pro Val  
 465 470 475 480  
 Phe Gly Gly Ile Leu Ser Leu Ile Asn Glu His Arg Ile Leu Ser Gly  
 485 490 495  
 Arg Pro Pro Leu Gly Phe Leu Asn Pro Arg Leu Tyr Gln Gln His Gly  
 500 505 510  
 Ala Gly Leu Phe Asp Val Thr Arg Gly Cys His Glu Ser Cys Leu Asp  
 515 520 525  
 Glu Glu Val Glu Gly Gln Gly Phe Cys Ser Gly Pro Gly Trp Asp Pro  
 530 535 540  
 Val Thr Gly Trp Gly Thr Pro Asn Phe Pro Ala Leu Leu Lys Thr Leu  
 545 550 555 560  
 Leu Asn Pro

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 587 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Lys Ser Ser Ala Ala Lys Gln Thr Val Leu Cys Leu Asn Arg Tyr  
 1 5 10 15  
 Ala Val Val Ala Leu Pro Leu Ala Ile Ala Ser Phe Ala Ala Phe Gly  
 20 25 30  
 Ala Ser Pro Ala Ser Thr Leu Trp Ala Pro Thr Asp Thr Lys Ala Phe  
 35 40 45  
 Val Thr Pro Ala Gln Val Glu Ala Arg Ser Ala Ala Pro Leu Leu Glu  
 50 55 60  
 Leu Ala Ala Gly Glu Thr Ala His Ile Val Val Ser Leu Lys Leu Arg  
 65 70 75 80  
 Asp Glu Ala Gln Leu Lys Gln Leu Ala Gln Ala Val Asn Gln Pro Gly  
 85 90 95

Asn Ala Gln Phe Gly Lys Phe Leu Lys Arg Arg Gln Phe Leu Ser Gln  
100 105 110

Phe Ala Pro Thr Glu Ala Gln Val Gln Ala Val Val Ala His Leu Arg  
115 120 125

Lys Asn Gly Phe Val Asn Ile His Val Val Pro Asn Arg Leu Leu Ile  
130 135 140

Ser Ala Asp Gly Ser Ala Gly Ala Val Lys Ala Ala Phe Asn Thr Pro  
145 150 155 160

Leu Val Arg Tyr Gln Leu Asn Gly Lys Ala Gly Tyr Ala Asn Thr Ala  
165 170 175

Pro Ala Gln Val Pro Gln Asp Leu Gly Glu Ile Val Gly Ser Val Leu  
180 185 190

Gly Leu Gln Asn Val Thr Arg Ala His Pro Met Leu Lys Val Gly Glu  
195 200 205

Arg Ser Ala Ala Lys Thr Leu Ala Ala Gly Thr Ala Lys Gly His Asn  
210 215 220

Pro Thr Glu Phe Pro Thr Ile Tyr Asp Ala Ser Ser Ala Pro Thr Ala  
225 230 235 240

Ala Asn Thr Thr Val Gly Ile Ile Thr Ile Gly Gly Val Ser Gln Thr  
245 250 255

Leu Gln Asp Leu Gln Gln Phe Thr Ser Ala Asn Gly Leu Ala Ser Val  
260 265 270

Asn Thr Gln Thr Ile Gln Thr Gly Ser Ser Asn Gly Asp Tyr Ser Asp  
275 280 285

Asp Gln Gln Gly Gln Gly Glu Trp Asp Leu Asp Ser Gln Ser Ile Val  
290 295 300

Gly Ser Ala Gly Gly Ala Val Gln Gln Leu Leu Phe Tyr Met Ala Asp  
305 310 315 320

Gln Ser Ala Ser Gly Asn Thr Gly Leu Thr Gln Ala Phe Asn Gln Ala  
325 330 335

Val Ser Asp Asn Val Ala Lys Val Ile Asn Val Ser Leu Gly Trp Cys  
340 345 350

Glu Ala Asp Ala Asn Ala Asp Gly Thr Leu Gln Ala Glu Asp Arg Ile  
355 360 365

Phe Ala Thr Ala Ala Ala Gln Gly Gln Thr Phe Ser Val Ser Ser Gly  
370 375 380

Asp Glu Gly Val Tyr Glu Cys Asn Asn Arg Gly Tyr Pro Asp Gly Ser  
385 390 395 400

Thr Tyr Ser Val Ser Trp Pro Ala Ser Ser Pro Asn Val Ile Ala Val  
405 410 415

Gly Gly Thr Thr Leu Tyr Thr Ser Ala Gly Ala Tyr Ser Asn Glu  
420 425 430

Thr Val Trp Asn Glu Gly Leu Asp Ser Asn Gly Lys Leu Trp Ala Thr  
 435 440 445  
 Gly Gly Gly Tyr Ser Val Tyr Glu Ser Lys Pro Ser Trp Gln Ser Val  
 450 455 460  
 Val Ser Gly Thr Pro Gly Arg Arg Leu Leu Pro Asp Ile Ser Phe Asp  
 465 470 475 480  
 Ala Ala Gln Gly Thr Gly Ala Leu Ile Tyr Asn Tyr Gly Gln Leu Gln  
 485 490 495  
 Gln Ile Gly Gly Thr Ser Leu Ala Ser Pro Ile Phe Val Gly Leu Trp  
 500 505 510  
 Ala Arg Leu Gln Ser Ala Asn Ser Asn Ser Leu Gly Phe Pro Ala Ala  
 515 520 525  
 Ser Phe Tyr Ser Ala Ile Ser Ser Thr Pro Ser Leu Val His Asp Val  
 530 535 540  
 Lys Ser Gly Asn Asn Gly Tyr Gly Gly Tyr Gly Tyr Asn Ala Gly Thr  
 545 550 555 560  
 Gly Trp Asp Tyr Pro Thr Gly Trp Gly Ser Leu Asp Ile Ala Lys Leu  
 565 570 575  
 Ser Ala Tyr Ile Arg Ser Asn Gly Phe Gly His  
 580 585

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 635 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Lys Ile Glu Lys Thr Ala Leu Thr Val Ala Ile Ala Leu Ala Met  
 1 5 10 15

Ser Ser Leu Ser Ala His Ala Glu Asp Ala Trp Val Ser Thr His Thr  
 20 25 30

Gln Ala Ala Met Ser Pro Pro Ala Ser Thr Gln Val Leu Ala Ala Ser  
 35 40 45

Ser Thr Ser Ala Thr Thr Thr Gly Asn Ala Tyr Thr Leu Asn Met Thr  
 50 55 60

Gly Ser Pro Arg Ile Asp Gly Ala Ala Val Thr Ala Leu Glu Ala Asp  
 65 70 75 80

His Pro Leu His Val Glu Val Ala Leu Lys Leu Arg Asn Pro Asp Ala

85

90

95

Leu Gln Thr Phe Leu Ala Gly Val Thr Thr Pro Gly Ser Ala Leu Phe  
 100 105 110  
 Gly Lys Phe Leu Thr Pro Ser Gln Phe Thr Glu Arg Phe Gly Pro Thr  
 115 120 125  
 Gln Ser Gln Val Asp Ala Val Val Ala His Leu Gln Gln Ala Gly Phe  
 130 135 140  
 Thr Asn Ile Glu Val Ala Pro Asn Arg Leu Leu Ile Ser Ala Asp Gly  
 145 150 155 160  
 Thr Ala Gly Ala Ala Thr Asn Gly Phe Arg Thr Ser Ile Lys Arg Phe  
 165 170 175  
 Ser Ala Asn Gly Arg Glu Phe Phe Ala Asn Asp Ala Pro Ala Leu Val  
 180 185 190  
 Pro Ala Ser Leu Gly Asp Ser Val Asn Ala Val Leu Gly Leu Gln Asn  
 195 200 205  
 Val Ser Val Lys His Thr Leu His His Val Tyr His Pro Glu Asp Val  
 210 215 220  
 Thr Val Pro Gly Pro Asn Val Gly Thr Gln Ala Ala Ala Ala Val Ala  
 225 230 235 240  
 Ala His His Pro Gln Asp Phe Ala Ala Ile Tyr Gly Gly Ser Ser Leu  
 245 250 255  
 Pro Ala Ala Thr Asn Thr Ala Val Gly Ile Ile Thr Trp Gly Ser Ile  
 260 265 270  
 Thr Gln Thr Val Thr Asp Leu Asn Ser Phe Thr Ser Gly Ala Gly Leu  
 275 280 285  
 Ala Thr Val Asn Ser Thr Ile Thr Lys Val Gly Ser Gly Thr Phe Ala  
 290 295 300  
 Asn Asp Pro Asp Ser Asn Gly Glu Trp Ser Leu Asp Ser Gln Asp Ile  
 305 310 315 320  
 Val Gly Ile Ala Gly Gly Val Lys Gln Leu Ile Phe Tyr Thr Ser Ala  
 325 330 335  
 Asn Gly Asp Ser Ser Ser Gly Ile Thr Asp Ala Gly Ile Thr Ala  
 340 345 350  
 Ser Tyr Asn Arg Ala Val Thr Asp Asn Ile Ala Lys Leu Ile Asn Val  
 355 360 365  
 Ser Leu Gly Glu Asp Glu Thr Ala Ala Gln Gln Ser Gly Thr Gln Ala  
 370 375 380  
 Ala Asp Asp Ala Ile Phe Gln Gln Ala Val Ala Gln Gly Gln Thr Phe  
 385 390 395 400  
 Ser Ile Ala Ser Gly Asp Ala Gly Val Tyr Gln Trp Ser Thr Asp Pro  
 405 410 415  
 Thr Ser Gly Ser Pro Gly Tyr Val Ala Asn Ser Ala Gly Thr Val Lys

420	425	430
Ile Asp Leu Thr His Tyr Ser Val Ser Glu Pro Ala Ser Ser Pro Tyr		
435	440	445
Val Ile Gln Val Gly Gly Thr Thr Leu Ser Thr Ser Gly Thr Thr Trp		
450	455	460
Ser Gly Glu Thr Val Trp Asn Glu Gly Leu Ser Ala Ile Ala Pro Ser		
465	470	475
Gln Gly Asp Asn Asn Gln Arg Leu Trp Ala Thr Gly Gly Val Ser		
485	490	495
Leu Tyr Glu Ala Ala Pro Ser Trp Gln Ser Ser Val Ser Ser Ser Thr		
500	505	510
Lys Arg Val Gly Pro Asp Leu Ala Phe Asp Ala Ala Ser Ser Ser Gly		
515	520	525
Ala Leu Ile Val Val Asn Gly Ser Thr Glu Gln Val Gly Gly Thr Ser		
530	535	540
Leu Ala Ser Pro Leu Phe Val Gly Ala Phe Ala Arg Ile Glu Ser Ala		
545	550	555
Ala Asn Asn Ala Ile Gly Phe Pro Ala Ser Lys Phe Tyr Gln Ala Phe		
565	570	575
Pro Thr Gln Thr Ser Leu Leu His Asp Val Thr Ser Gly Asn Asn Gly		
580	585	590
Tyr Gln Ser His Gly Tyr Thr Ala Ala Thr Gly Phe Asp Glu Ala Thr		
595	600	605
Gly Phe Gly Ser Phe Asp Ile Gly Lys Leu Asn Thr Tyr Ala Gln Ala		
610	615	620
Asn Trp Val Thr Gly Gly Gly Ser Thr		
625	630	635

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Oligonucleotides"
- (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GTGATCACAG AATGGCACTT

20

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Oligonucleotides"
- (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AACATGGGTT TCCGTAGGTC

20

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Oligonucleotides"
- (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CTTCCTCAGG GTCCGCACGG

20

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 38 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Oligonucleotides"
- (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TGTAAAACGA CGGCCAGTCA GACCTTCCAG TAGGGACC

38

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 38 base pairs
  - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "Oligonucleotides"  
(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CAGGAAACAG CTATGACCCT GTATCCCACA CAAGAGAT

38

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 38 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "Oligonucleotides"  
(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGTAAAACGA CGGCCAGTTA GATGCCATTG GGGACTGG

38

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 38 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "Oligonucleotides"  
(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CAGGAAACAG CTATGACCGT CATGGAAATA CTGCTCCA

38